Data Model and Management of Plant Genetic Resources

Masamichi UMEHARA*

Abstract

Data management system of the NIAR gene bank was developed on the basis of a relation­
dional data model using relational data base management system (RDBMS) which is running
under the UNIX operating system. Stored data are classified into three categories: "pass­
port", "stock", "evaluation"data. A total of 170,000 records has been compiled for passport
data, along with 300,000 stock data and 100,000 evaluation data, so far.

Data model of plant genetic resources is as follows:

The schema for the passport data management consists of 12 tables, e. g. a passport
data table, a plant code table, etc. The schema for the stock data management consists of 14 tables, e. g. tables related to
the registration of new accessions, distribution of seeds, multiplication of seeds, germination
test of seeds, etc.

The schema for the management of evaluated data consists of more than 600 tables.

Evaluation was performed for approximately 110 plants.

As usually data management of one table requires a data management program, the
passport data management system and the stock data management system were readily
developed, under such concept.

For the management of evaluated data, however; it is impossible to develop 600 pro­
grams for the management of 600 tables. Therefore, instead of the development of specific
programs for each plant, we developed a data dictionary system relating to the table struc­
ture for each plant and also a data management program which covers the whole tables for
the plants using the dynamic SQL.

Outline of schema

A data model for the plant genetic resources stored at NIAR has been developed on the basis of a re­
lational data model. This schema consists of about 700 tables, which can be classified into 3 categories
(Fig. 1).

In the first category, passport data are the basis of the plant genetic resources database. For the
identification of individual genetic resources, the passport data are firstly constructed at the time of col­
lection or reception of the accessions. Every genetic resource has it's own accession number which is
never duplicated among the plant genetic resources of MAFF.

In the second category, the stock control data are very important for the management of plant ge­
genetic resources, because the distribution of seeds upon request by researchers is a major working. In this
management, data such as germination rates, seed weight, addresses in the preservation room are essen­
tial. Each plant genetic resources is identified only by the accession number.

In the third category, data of the evaluated characteristics are most useful for users of genetic re­
sources. Table structures of evaluation data may/ should be different for each species. Then, a large

Presented at the 27th International Symposium on "Plant Genetic Resource Management in the
Tropics", Tsukuba, Ibaraki, Japan, 25-26 August 1993, held by Tropical Agriculture Research Center
(TARC).

*Department of Genetic Resources I, National Institute of Agrobiological Resources, Tsukuba, Ibaraki
305 Japan
number of tables, presently over 600 tables, is required. The management of evaluated data is, therefore, the most complex procedure.

**Passport data**

The basic structure of the passport data is shown in Fig. 2. For the purpose of normalization, every row of passport data contains an accession number, an institute code, a laboratory code, a plant code and an origin code. As a result of normalization, the passport data table refers to 4 foreign tables, i.e. an institution table, a laboratory table, a plant table and a country table. Passport data are subdivided and stored in 3 tables of the same structure, "pass", "pass_sub" and "pass_wk" based on actual efficiency of data management. Some other additional tables are also used for the data management such as determination of accession number. The accession numbers of passport data are, then, used for the stock control data and for the character evaluation data.

**Stock control data**

"Address", a table for storage addresses of seed bottles in the preservation room, has columns for accession numbers, addresses, germination rates, amount of seeds, warehousing days, etc.

Two tables are used for the distribution, "request_no" and "request_item". The table "request_no" is used for the attributes of user who requested and "request_item" is used for the accessions requested.

For the multiplication and for the test of germination rates, tables of a similar structure are used.

Fig. 3 shows tables for the distribution.

**Evaluation data**

All the evaluation data commonly have an accession number, an institution code, a laboratory code in which the genetic resource was evaluated, and the year of evaluation. Other attributes vary depending on the plant groups. All the attributes of characteristics which are evaluated for each plant are included in a data dictionary system. The number of tables for the data of evaluated characteristics are over 600, and those tables are created automatically by a program (Fig. 4).
Fig. 2 Passport data

**Passport** (4 tables with similar structure)
(pass, pass_sub, pass_wk, pass_pref)
Table of passport data for individual varieties
- accession number
- date of registration
- institute code
- laboratory code
- plant code
- variety name
- origin code
- source code
- method of storage

**Institution**
- institute code
- institute name
- abbreviation
- address code

**Laboratory**
Table of laboratories
- institute code
- laboratory code
- division name
- laboratory name
- abbreviation
- address code

**Inst_address**
Table of address of institute
- institute code
- address
- phone number
- address code

**Plant**
Table of basic data of species
- plant code
- plant group code
- scientific name
- plant name

**Plant_group**
Table of basic data for plant group
- plant group code
- group name

**Country_code**
Table of areas
- country code (same for origin and source)
- name of nation (or prefecture)
Fig. 3 Tables for the distribution

Evaluation data tables

Fig. 4 Evaluation data
Data processor in the genebank, NIAR

Fig. 5 shows the concept of data processing in the genebank. Application programs for data management are executed on a UNIX operating system. In the database server, 10 to 20 processes are normally running for the database access.

Data management

A data management system has been developed for the passport data and the stock control data by an ordinary method.

The stock control management consists of 4 sub-systems, as follows:
1) New seed reception
2) Distribution of genetic resources
3) Multiplication of genetic resources
4) Germination testing

Ability for OLTP (OnLine Transaction Processing) of DBMS is required for the management of the stock control data.

Data management for the evaluated characteristics

By using ordinary methods, the development of a data management system for evaluated characteristics is difficult, because too many tables are required. Instead of using ordinary methods, we introduced a data dictionary system.

In the first step, we developed a management system for the data dictionary, then we input whole data which define the characteristics for the evaluation of each plant.

In the next step, a program which generates SQL statement for creating tables was developed, by us-

Fig. 5 Outline of data processing in genebank, MAFF
ing the data dictionary. Fig. 6 shows SQL statements generated.

In the last step, a program which can handle the whole evaluation table by using the data dictionary was developed. Operators can select any table of evaluation data with a menu system. After the decision of the plant, this program selects whole definition data from data dictionary for the plant, then allocates necessary memories and manages the data of the table by the dynamic SQL (Fig. 7).

Fig. 8-Fig. 11 show examples of screen images.

```
CREATE TABLE t02001 (h1 (
    acce_no char (8) NOT NULL,
    inst_code char (5) NOT NULL,
    lab_code char (5) NOT NULL,
    year char (4) NOT NULL check (year > "1970"),
    h1_001 char (1) check (h1_001 between "2" and "8"),
    h1_002 integer CHECK (h1_002 > 0 AND h1_002 < 9999),
    h1_003 decimal (5, 1) CHECK (h1_003 > 0 AND h1_003 < 999.9),
    h1_004 char (1) check (h1_004 in ("0", "2", "3", "4", "5", "6", "7", "8")),
    h1_005 char (1) check (h1_005 between "1" and "9"),
    h1_006 char (1) check (h1_006 between "2" and "8"),
    h1_007 char (1) check (h1_007 between "0" and "9"),
    h1_008 char (5),
    h1_009 char (5),
    foreign key (inst_code) REFERENCES inst_code,
    foreign key (inst_code, lab_code) REFERENCES lab_code,
    Primary key (acce_no, inst_code, lab_code, year)
);  
```

Fig. 6 Example of SQL statements generated

Fig. 7 Data management for the evaluated characteristics
Fig. 8 Top menu

特性データ管理システム

<table>
<thead>
<tr>
<th>No.</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>07011 にほんなし</td>
<td>11</td>
</tr>
<tr>
<td>2</td>
<td>07002 せいようなし</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>07003 マルメロ</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>07004 おとう</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>07005 きいちご</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>07006 カーランツ</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>07007 グーテベリー</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>07008 ブルーベリー</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>07009 くるみ</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>07010 くり</td>
<td>20</td>
<td></td>
</tr>
</tbody>
</table>

選択して下さい（終了 E）

Fig. 9 Menu for fruits

特性データ管理システム

<table>
<thead>
<tr>
<th>No.</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>02</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>04</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>05</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>9</td>
<td>06</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>10</td>
<td>07</td>
<td>7</td>
</tr>
<tr>
<td>7</td>
<td>11</td>
<td>08</td>
<td>8</td>
</tr>
<tr>
<td>8</td>
<td>12</td>
<td>09</td>
<td>9</td>
</tr>
<tr>
<td>9</td>
<td>13</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

選択して下さい（終了 E）
Discussion

Riley, K. (IBPGR): Can the documentation system of NIAR be utilized in gene banks of other countries?
Answer: Yes but with some modifications.

Riley, K. (IBPGR): Comment: The challenge is both to develop powerful computer systems that can manage the huge information needed to keep track of nucleotide sequences in genome of plants as well as to develop information systems appropriate for PGR management in smaller national programs with limited resources.